Coursera- IBM Data Science Professional Certification Program
Applied Data Science
Capstone Course

Project: Battle of Neighborhoods – New Restaurant in Central Ohio, US

By Venkatachalapathy Punugu

Feb 11, 2021

Table of Contents

1. Introduction

A. Business Problem

Ohio is a state in the East North Central Region of the Midwestern United States – it is the 17th state and was incorporated into Union in 1803. Ohio has given **8 Presidents (Presidents - WH Harrison, Grant, Hayes, Garfield, B Harrison, McKinley, Taft, Harding)**, notable astronauts like **John Glenn** (first American in space), **Neil Armstrong** (first man to walk on moon – Apollo 11), **John Lovell** (famous for bringing back the crew of Apollo 13 alive), **Judith Resnick** (part of space shuttle Challenger crew exploded shortly after takeoff). Franklin County (named after **Benjamin Franklin**, statesman, scientist, and inventor) is in central Ohio and is the jurisdiction which houses capitol city of Columbus which is the seat of power for the State of Ohio. Columbus is also the largest city in Ohio.

With its business-friendly atmosphere, large vibrant population, growing diverse demography, Columbus and its immediate neighborhood cities in Franklin County offer many business opportunities for new businesses. Franklin County also houses many of the Top 20 cities in terms of Median Family Income, one of the important indicators of disposable income which will influence the spend on the consumption especially on food and restaurants. It is also observed that Asian descent and especially Indian descent population is one of the top earning demography and therefore an opportunity for the success of the business with a focus on their culture and tastes.

The present exercise is to study Columbus and its neighborhood of Franklin County with a view to recommend three cities for establishing an Indian Restaurant in one of them.

B. Target Audience:

a. Entrepreneurs targeting to open Indian Restaurant in the state.

b. Any other interested party or established entities in restaurant business in opening a restaurant in Central Ohio.

c. Data Science Students to carry forward the analysis and bring out new insights with the help of additional data/analysis/research.

2. Data

a. **City Data:** City of Columbus lies within the Franklin County and is surrounded by several cities in suburbs. These cities also fall within the jurisdiction of Franklin County. Some of these cities are also target for the location of the restaurant. Therefore, the list of cities in the Franklin County is used for the analysis.

b. **Location**: **GEOPY** is used to retrieve the latitude and longitude data and we will be using the geocode function for the purpose.

c. **Venues in the cities**: **FOURSQUARE API** is used to retrieve all the venues within the cities in the analysis

d. A radius of 500 meters were used to retrieve the venues in each of the cities – these however were giving a very short list and at the same time, based on the latitude and longitude of the center of the target city, resulted in not picking up the venues then acc and city. Some important cities like Dublin were totally missing. Therefore, the radius was increased to **2000 meters**. This provided a good selection of the venues across all the cities in the county.

e. **Income Data**: We will be using a commercial data that is derived from the US Census to retrieve the median income of the cities.

f. **SKLEARN** – Scikit Learn package is used for arriving at the KMean for the venues.

g. **Visualization**:
   a. **Folium** package is used for the geographical mapping and cluster visualization.
   b. **Yellowbrick** is used for the visualization using the KElbowVisualizer for displaying the optimum number of K-Mean clusters to be used for grouping the venues.
   c. **Matplotlib.pyplot** is used for the visualization of the venues across the cities

h. **Data Cleanup**: As needed based on the availability of the above data.

i. **Data Usage:** The above data will be used to arrive at top 3 cities for the recommendation.

3. Methodology

The first step is to get the data set up, explored and cleaned up. This is followed by the analysis of the venues in all the cities in the county.

A. Data Setup/Exploration/Cleanup/Wrangling

1. For the analysis of the data, the data from the state has been taken from **Wikipedia**.

2. This data was then used to extract at the City level for all cities in Franklin County – there are total of **16 cities** in the county.

3. The city center **Latitude** and **Longitude** location information were retrieved form **GEOPY** service using the standard API – *geocode*. Since Franklin County had a few cities (4 Cities) that were having the same name as other international cities, state (OH) was used to qualify the city.

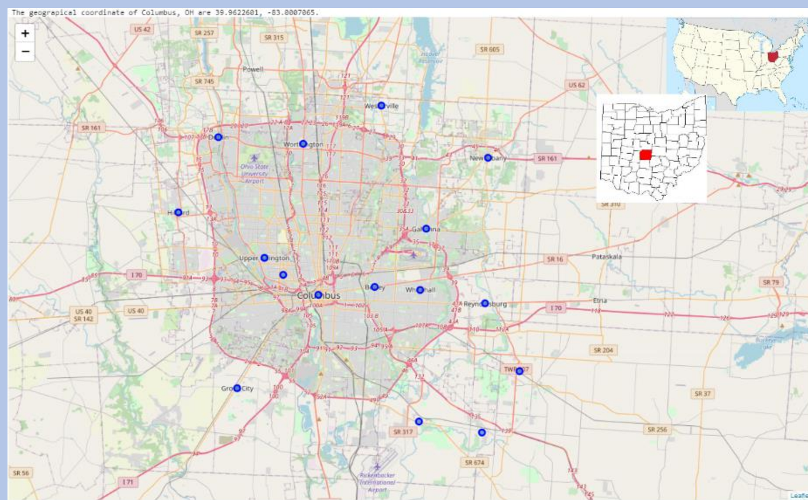
4. Then the venues in these cities around the city centers [given by the Latitude and Longitude] are retrieved from **FOURSQUARE** using the **/V2/Venues/Explore** API. These are regular APIs. The radius of **2000 meters** is used to get a reasonable set of venues – as anything less gave a small set of venues.

5. For visualization of the data, **matplotlib.pyplot** is used for the bar charts and **Folium** is used for the display of the locations on the map.

6. Population and Income data is retrieved from the **Cleveland.com** – a premier media outlet in Ohio. The data that is being used is cured by Cleveland.com from US Government Census. We will be using the **Population** and **Median Family Income** for all the cities in the Franklin County for our analysis.

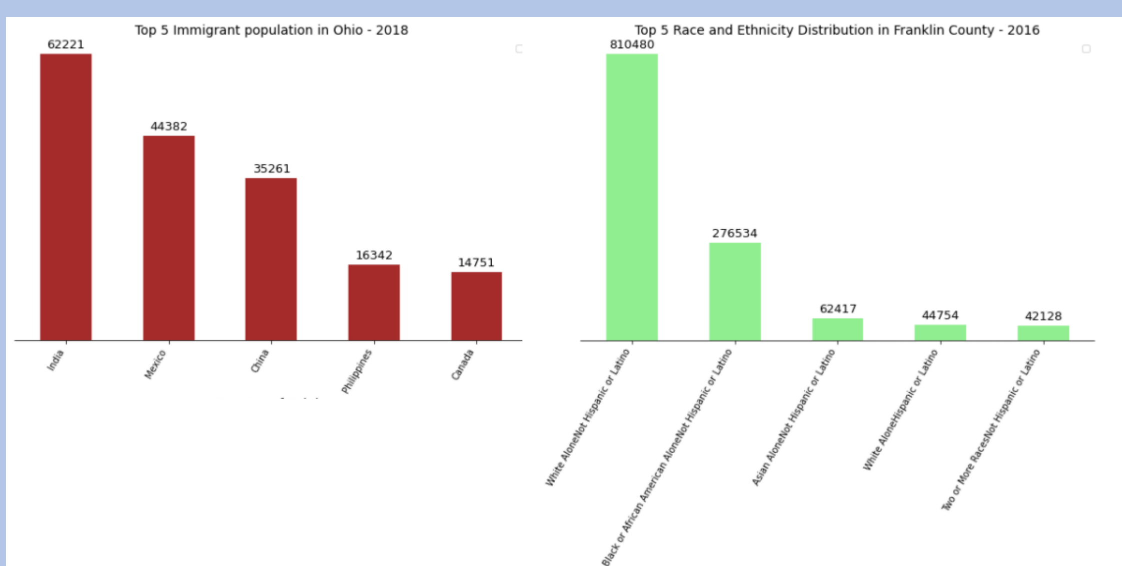
B. Analysis

1. **Where is Franklin County in US?** Let us take a look at the Franklin County with all the cities identified – we use the Folium Map of Franklin County with 16 cities:



The US and State of Ohio Map is provided as a guidance for the subject location for those international audience, not familiar with the geography, who may also be interested. Also notice how well the county and the cities in the county have related to freeways. This will also play a role in the analysis and conclusion.

2. **Who constitutes the county?** Let us look at the population data – later in the analysis we will also look at the population in each of the 16 cities.



```python
xnurl_data = pd.read_csv('Global Diversity.csv')
data2018=xnurl_data[xnurl_data['Year']==2018].sort_values(by='Total Population', ascending=False)
top5=data2018.head(5)
top5
```

[48]:

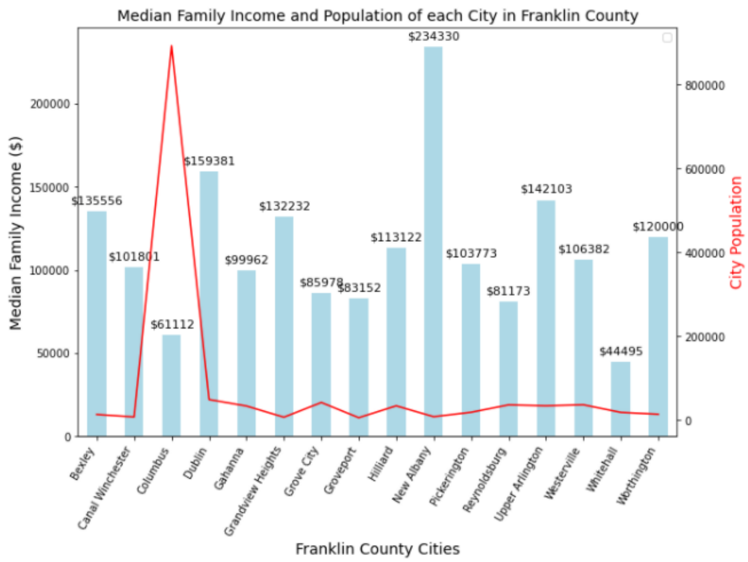
	ID Birthplace	Birthplace	ID Year	Year	ID Nativity	Nativity	Country Code	Total Population	Total Population MOE Appx	Geography	ID Geography
51	210	India	2018	2018	2	Foreign born	IND	62221	6107.793218	Ohio	04000US39
82	303	Mexico	2018	2018	2	Foreign born	MEX	44382	5162.406135	Ohio	04000US39
49	207	China	2018	2018	2	Foreign born	CHN	35261	4603.265881	Ohio	04000US39
67	233	Philippines	2018	2018	2	Foreign born	PHL	16342	3136.343160	Ohio	04000US39
81	301	Canada	2018	2018	2	Foreign born	CAN	14751	2979.965919	Ohio	04000US39

```python
xnurl_data = pd.read_csv('Race and Ethnicity.csv')
data2016=xnurl_data[xnurl_data['Year']==2016].sort_values(by='Population', ascending=False)
top5=data2016.head(5)
top5
```

[45]:

	ID Race	Race	ID Ethnicity	Ethnicity	ID Year	Year	Hispanic Population Moe	Geography	ID Geography	Slug Geography	Population	share
28	0	White Alone	0	Not Hispanic or Latino	2016	2016	1193.000000	Franklin County, OH	05000US39049	franklin-county-oh	810480	0.640940
30	1	Black or African American Alone	0	Not Hispanic or Latino	2016	2016	4459.000000	Franklin County, OH	05000US39049	franklin-county-oh	276534	0.218687
34	3	Asian Alone	0	Not Hispanic or Latino	2016	2016	1867.000000	Franklin County, OH	05000US39049	franklin-county-oh	62417	0.049360
29	0	White Alone	1	Hispanic or Latino	2016	2016	4662.000000	Franklin County, OH	05000US39049	franklin-county-oh	44754	0.035392
40	6	Two or More Races	0	Not Hispanic or Latino	2016	2016	5056.380128	Franklin County, OH	05000US39049	franklin-county-oh	42128	0.033315

3. **Looking at Cities – Population and Median Family Income:** The income data from the Census points to the affluent neighborhoods in the Franklin county shows that the high-income families reside in these cities – New Albany, Dublin, Upper Arlington, Bexley, Grandview Heights, Worthington, Hilliard. The rest of the cities follow. This is an important criterion to keep in mind.



Rankfamilyincome	City	Population	Medianfamilyincome
2	New Albany	7724	$234,330
5	Dublin	48647	$159,381
8	Upper Arlington	33771	$142,103
10	Bexley	13057	$135,556
11	Grandview Heights	6536	$132,232
22	Worthington	13575	$120,000
26	Hilliard	33672	$113,122
34	Westerville	36120	$106,382
37	Pickerington	18291	$103,773
40	Canal Winchester	7101	$101,801
41	Gahanna	33248	$99,962
62	Grove City	41627	$85,978
72	Groveport	5363	$83,152
81	Reynoldsburg	35893	$81,173
149	Columbus	892672	$61,112
233	Whitehall	18062	$44,495

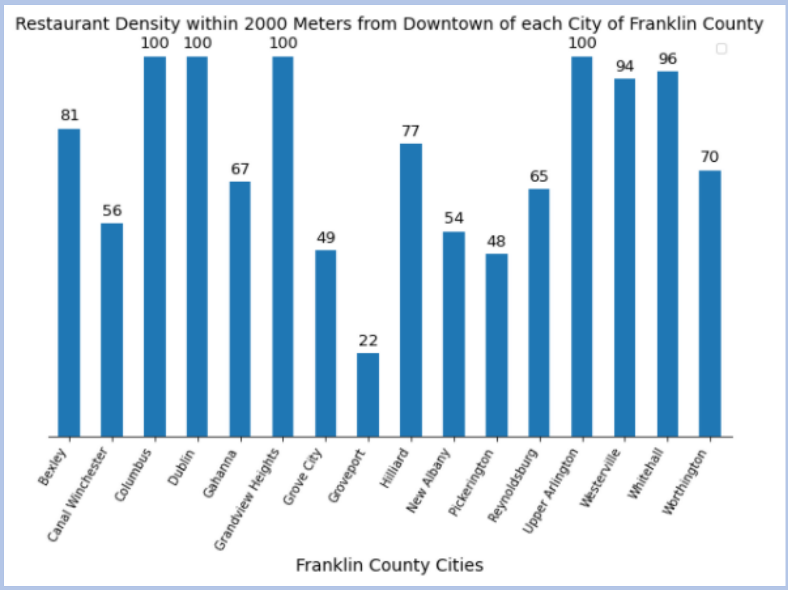
4. **Understanding more of the current scenarios – How many venues are in Franklin county in the dataset?** The dataset returned 1179 venues of all categories

```
[24]: print(franklin_cities_venues.shape)
      franklin_cities_venues.head()

      (1179, 7)
```

[24]:

	City	Latitude	Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Bexley	39.969238	-82.936864	Jeffrey Park	39.972572	-82.943011	Park
1	Bexley	39.969238	-82.936864	Franklin Park Conservatory and Botanical Gardens	39.965933	-82.952814	Garden
2	Bexley	39.969238	-82.936864	Giant Eagle Market District Express	39.957479	-82.939145	Supermarket
3	Bexley	39.969238	-82.936864	Franklin Park Community Garden Campus	39.967486	-82.951224	Garden
4	Bexley	39.969238	-82.936864	Giuseppe's Ritrovo	39.957316	-82.938282	Italian Restaurant

5. **What are the total Number of venues in each of the cities?** Using a 2000 meters radius around the center of each city has provided us with a good number of venues for use in our analysis.



6. **Drilling down a little more – how many categories of the venues?** Since the venue data comes with 'Venue Category' field, to make a proper comparison and to evaluate means, these are converted into appropriate numerical fields/columns using the Pandas function get_dummies for the dataframe. This gave rise to total of 205 unique features in the venues data.

```
[26]: print('There are {} uniques categories.'.format(len(franklin_cities_venues['Venue Category'].unique())))

      There are 205 uniques categories.
```

7. **Further let us see which are the more popular locations in the cities:** This will provide an insight into the people's preference of spending time. Barring a few instances of

cities, across all the other cities, it is found that the top three most common venue is a food eatery.

[45]:

	City	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
0	Bexley	Pizza Place	Bank	Sandwich Place
1	Canal Winchester	Pizza Place	American Restaurant	Bank
2	Columbus	Coffee Shop	Park	Bar
3	Dublin	Hotel	Pizza Place	Italian Restaurant
4	Gahanna	Sandwich Place	Pizza Place	American Restaurant
5	Grandview Heights	Coffee Shop	Bar	Pizza Place
6	Grove City	Pizza Place	Sandwich Place	Ice Cream Shop
7	Groveport	Pizza Place	Park	Bar
8	Hilliard	Pizza Place	Fast Food Restaurant	Bank
9	New Albany	Coffee Shop	Bank	Gym / Fitness Center
10	Pickerington	Pizza Place	Pharmacy	Sandwich Place
11	Reynoldsburg	Pizza Place	Fast Food Restaurant	Bar
12	Upper Arlington	Coffee Shop	Mexican Restaurant	Italian Restaurant
13	Westerville	Pizza Place	Bank	Coffee Shop
14	Whitehall	Fast Food Restaurant	Gas Station	Discount Store
15	Worthington	Pizza Place	Italian Restaurant	American Restaurant

8. **Now let us see how the competition stacks up.** Let us now see how many 'Indian Restaurants' are in the data set. From the data set of the venues collected within the 2000 meters radius of the center of all cities in Franklin County, we see there are only 4 present in 4 cities. While there may be more outside this radius, there are not present in the dataset available for this exercise.

```
[86]: df_restaurants[df_restaurants['Venue Category'].str.contains('Indian')]
```

[86]:

	City	Latitude	Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
57	Bexley	39.969238	-82.936864	Aab India Restaurant	39.957236	-82.934840	Indian Restaurant
228	Columbus	39.962260	-83.000707	Indian Oven	39.957369	-82.987884	Indian Restaurant
422	Grandview Heights	39.979786	-83.040740	Aab India Restaurant	39.987737	-83.044456	Indian Restaurant
839	Upper Arlington	39.994508	-83.062408	Aab India Restaurant	39.987737	-83.044456	Indian Restaurant

It appears the for a county of more than a million population there are just handful of Indian Restaurants. So far so good!

9. **Deeper dive analysis:** Before we investigate the cities, let us first try to group them to identify the similarity among them. To do this we resort to the unsupervised learning Machine Learning (ML) Algorithm to group them. Clustering is the best ML approach to

analyze the dataset as we have categories of the venues to deal with. Before we analyze the clusters, we need to convert the categorical information into a numeric feature for the ML algorithm to process. We convert the categorical data into numerical by using **One Hot Encoding** – this will convert the column that we are interested in ('Venue Category') to numeric (0/1) – however this will increase the number of columns – one for each category – **206** total number columns/features:

```
[42]: # one hot encoding
franklin_cities_onehot = pd.get_dummies(franklin_cities_venues[['Venue Category']], prefix="", prefix_sep="")

# add neighborhood column back to dataframe
franklin_cities_onehot['City'] = franklin_cities_venues['City']

# move neighborhood column to the first column
fixed_columns = [franklin_cities_onehot.columns[-1]] + list(franklin_cities_onehot.columns[:-1])
franklin_cities_onehot = franklin_cities_onehot[fixed_columns]
print ("franklin_cities_onehot ", franklin_cities_onehot.shape)
franklin_cities_onehot.head()
```

```
franklin_cities_onehot  (1179, 206)
```

[42]:

	City	ATM	American Restaurant	Arepa Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	Automotive Shop	BBQ Joint	Bagel Shop	Bakery	Bank	Bar	Baseball Field	Baseball Stadium	Basketball Court
0	Bexley	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	Bexley	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	Bexley	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	Bexley	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	Bexley	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

10. **Grouping for cluster comparison:** Just so we compare all with the same scaling, we will take the means of each of them and group them by Cities.

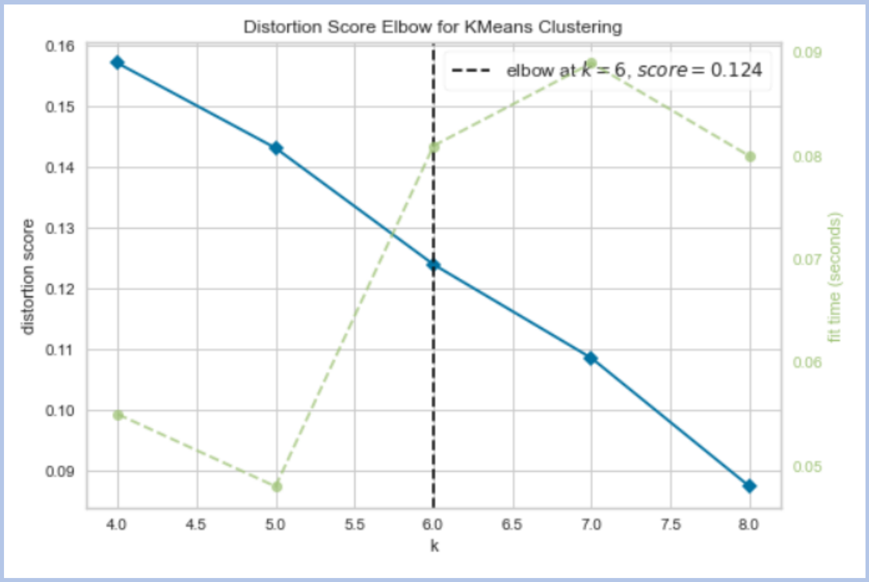
```
[43]: franklin_cities_grouped = franklin_cities_onehot.groupby('City').mean().reset_index()
print ("franklin_cities_grouped ", franklin_cities_grouped.shape)
franklin_cities_grouped
```

```
franklin_cities_grouped  (16, 206)
```

[43]:

	City	ATM	American Restaurant	Arepa Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	Automotive Shop	BBQ Joint	Bagel Shop	Bakery	Bank	Bar	Baseball Field	Baseball Stadium	Basketball Court	Bavarian Restaurant	Beer Store
0	Bexley	0.000000	0.024691	0.000000	0.000000	0.00	0.00	0.000000	0.012346	0.012346	0.000000	0.012346	0.037037	0.061728	0.000000	0.012346	0.000000	0.012346	0.00	0.000000
1	Canal Winchester	0.000000	0.089286	0.000000	0.000000	0.00	0.00	0.017857	0.000000	0.017857	0.000000	0.000000	0.035714	0.071429	0.017857	0.000000	0.000000	0.000000	0.00	0.000000
2	Columbus	0.000000	0.040000	0.000000	0.020000	0.01	0.01	0.000000	0.000000	0.000000	0.010000	0.000000	0.020000	0.000000	0.050000	0.000000	0.010000	0.000000	0.00	0.000000
3	Dublin	0.010000	0.010000	0.000000	0.000000	0.00	0.00	0.000000	0.000000	0.010000	0.000000	0.010000	0.000000	0.040000	0.040000	0.000000	0.000000	0.000000	0.00	0.000000
4	Gahanna	0.014925	0.059701	0.014925	0.000000	0.00	0.00	0.000000	0.000000	0.000000	0.014925	0.000000	0.000000	0.044776	0.044776	0.000000	0.000000	0.000000	0.00	0.000000
5	Grandview Heights	0.000000	0.030000	0.000000	0.020000	0.00	0.00	0.010000	0.010000	0.000000	0.010000	0.000000	0.020000	0.000000	0.060000	0.000000	0.000000	0.000000	0.01	0.000000
6	Grove City	0.020408	0.000000	0.000000	0.000000	0.00	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.020408	0.000000	0.020408	0.020408	0.020408	0.020408	0.00	0.000000
7	Groveport	0.000000	0.000000	0.000000	0.000000	0.00	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.045455	0.090909	0.000000	0.000000	0.000000	0.00	0.000000
8	Hilliard	0.000000	0.012987	0.000000	0.000000	0.00	0.00	0.012987	0.000000	0.000000	0.000000	0.000000	0.000000	0.051948	0.038961	0.012987	0.000000	0.000000	0.00	0.000000
9	New Albany	0.018519	0.037037	0.000000	0.018519	0.00	0.00	0.000000	0.000000	0.018519	0.000000	0.000000	0.000000	0.074074	0.000000	0.000000	0.000000	0.018519	0.00	0.000000
10	Pickerington	0.000000	0.000000	0.000000	0.000000	0.00	0.00	0.000000	0.000000	0.041667	0.000000	0.000000	0.000000	0.041667	0.020833	0.000000	0.000000	0.000000	0.00	0.000000
11	Reynoldsburg	0.000000	0.015385	0.000000	0.000000	0.00	0.00	0.000000	0.000000	0.015385	0.015385	0.000000	0.000000	0.046154	0.061538	0.000000	0.000000	0.000000	0.00	0.000000
12	Upper Arlington	0.000000	0.020000	0.000000	0.000000	0.00	0.00	0.000000	0.010000	0.000000	0.000000	0.010000	0.000000	0.020000	0.010000	0.000000	0.000000	0.000000	0.00	0.000000
13	Westerville	0.000000	0.031915	0.000000	0.000000	0.00	0.00	0.000000	0.010638	0.021277	0.000000	0.000000	0.031915	0.042553	0.000000	0.021277	0.000000	0.000000	0.00	0.000000
14	Whitehall	0.020833	0.010417	0.000000	0.000000	0.00	0.00	0.000000	0.000000	0.020833	0.000000	0.000000	0.000000	0.041667	0.000000	0.000000	0.000000	0.000000	0.00	0.010417
15	Worthington	0.014286	0.042857	0.000000	0.014286	0.00	0.00	0.000000	0.000000	0.000000	0.000000	0.014286	0.042857	0.042857	0.014286	0.000000	0.000000	0.000000	0.00	0.000000

11. **Optimal Clusters determination and Visualization:** Prior to clustering the data, let us find out how many clusters will be optimal for the data that we have. This is accomplished by using the Scikit package - Using the **SKLEARN algorithm of KMeans**, we find that the optimal number of clusters in the data appears to be 6. The **Yellowbrick's KElbowVisualizer** function provides us with the visualization of the optimal Number of Clusters (K=6) in the data:



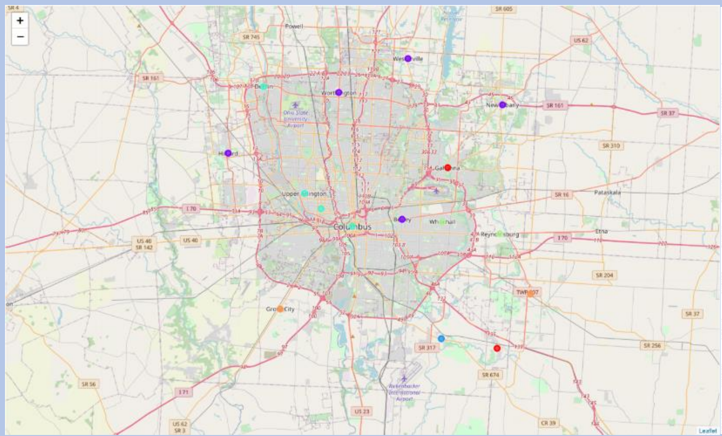
12. **Final List of Clusters:** We are now ready to look at the clusters that the ML Clustering KMeans algorithm has given - for this we perform the clustering first on the Means and later join them with the Cities and Venue Categories to identify the top 10 common venues in each of the cities.

```
[56]: # add clustering labels
#franklin_cities_venues_sorted.drop(['Cluster Labels'], axis=1, inplace=True)
franklin_cities_venues_sorted.insert(0, 'Cluster Labels', kmeans.labels_)

franklin_cities_merged = franklin_cities

franklin_cities_merged = franklin_cities_merged.join(franklin_cities_venues_sorted.set_index('City'), on='City')
print("franklin_cities_merged: ", franklin_cities_merged.shape)
franklin_cities_merged # check the last columns!

franklin_cities_merged:  (16, 14)
```

[56]:

	City	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
23	Bexley	39.9692	-82.9369	1	Pizza Place	Bank	Sandwich Place	Bakery	Coffee Shop	Park	Ice Cream Shop	Mediterranean Restaurant	Garden	Italian Restaurant
38	Canal Winchester	39.8429	-82.8159	0	Pizza Place	American Restaurant	Bank	Sandwich Place	Fast Food Restaurant	Mexican Restaurant	Bakery	Coffee Shop	Mobile Phone Shop	Gas Station
58	Columbus	39.9623	-83.0007	3	Coffee Shop	Park	Bar	American Restaurant	Brewery	Restaurant	Café	Hotel	Taco Place	Italian Restaurant
71	Dublin	40.0992	-83.1141	3	Hotel	Pizza Place	Italian Restaurant	Bank	Bar	Café	Sandwich Place	Park	Department Store	Ice Cream Shop
91	Gahanna	40.0196	-82.8791	0	Sandwich Place	Pizza Place	American Restaurant	Bar	Mexican Restaurant	Discount Store	Ice Cream Shop	Bank	Park	Chinese Restaurant
97	Grandview Heights	39.9798	-83.0407	3	Coffee Shop	Bar	Pizza Place	American Restaurant	Wine Shop	Sandwich Place	Italian Restaurant	Mexican Restaurant	Hotel	Grocery Store
100	Grove City	39.8815	-83.093	5	Pizza Place	Sandwich Place	Ice Cream Shop	Park	Soccer Field	Diner	Discount Store	Video Game Store	Fast Food Restaurant	Gas Station
101	Groveport	39.8522	-82.8869	2	Pizza Place	Park	Bar	Bank	Soccer Field	Smoke Shop	Chinese Restaurant	Supermarket	Mexican Restaurant	Coffee Shop
106	Hilliard	40.0338	-83.1596	1	Pizza Place	Fast Food Restaurant	Bank	Ice Cream Shop	Gym / Fitness Center	Pharmacy	Trail	Sandwich Place	Bar	Mexican Restaurant
161	New Albany	40.0812	-82.8088	1	Coffee Shop	Bank	Gym / Fitness Center	Pizza Place	Sandwich Place	Supermarket	Chinese Restaurant	American Restaurant	Hotel	Gym
191	Pickerington	39.8963	-82.7732	5	Pizza Place	Pharmacy	Sandwich Place	Coffee Shop	Automotive Shop	Bank	Video Store	Supermarket	Gas Station	Pet Store
199	Reynoldsburg	39.9548	-82.8121	4	Pizza Place	Fast Food Restaurant	Bar	Park	Bank	Discount Store	Mexican Restaurant	Gas Station	Chinese Restaurant	Grocery Store
244	Upper Arlington	39.9945	-83.0624	3	Coffee Shop	Mexican Restaurant	Italian Restaurant	Salon / Barbershop	Sandwich Place	Pizza Place	Wine Shop	Bank	Clothing Store	Wings Joint
261	Westerville	40.1261	-82.9295	1	Pizza Place	Bank	Coffee Shop	American Restaurant	Breakfast Spot	Fast Food Restaurant	Park	Bakery	Thrift / Vintage Store	Sandwich Place
263	Whitehall	39.9667	-82.8855	4	Fast Food Restaurant	Gas Station	Discount Store	Pizza Place	Fried Chicken Joint	Bank	Cosmetics Shop	Pharmacy	Sandwich Place	Chinese Restaurant
271	Worthington	40.0931	-83.018	1	Pizza Place	Italian Restaurant	American Restaurant	Bank	Bakery	Ice Cream Shop	Salon / Barbershop	Sandwich Place	Coffee Shop	Gym / Fitness Center

13. **Visualization of the clusters with the help of the Folium Map.** This completed the clustering the dataset, let us see how it looks on the map:



Cluster Labels	level_0	index	City	Latitude	Longitude	1st Most Common Venue
0	1	1	Canal Winchester	39.8429	-82.8159	Pizza Place
0	4	4	Gahanna	40.0196	-82.8791	Sandwich Place
1	0	0	Bexley	39.9692	-82.9369	Pizza Place
1	8	8	Hilliard	40.0338	-83.1596	Pizza Place
1	9	9	New Albany	40.0812	-82.8088	Coffee Shop
1	13	13	Westerville	40.1261	-82.9295	Pizza Place
1	15	15	Worthington	40.0931	-83.018	Pizza Place
2	7	7	Groveport	39.8522	-82.8869	Pizza Place
3	2	2	Columbus	39.9623	-83.0007	Coffee Shop
3	3	3	Dublin	40.0992	-83.1141	Hotel
3	5	5	Grandview Heights	39.9798	-83.0407	Coffee Shop
3	12	12	Upper Arlington	39.9945	-83.0624	Coffee Shop
4	11	11	Reynoldsburg	39.9548	-82.8121	Pizza Place
4	14	14	Whitehall	39.9667	-82.8855	Fast Food Restaurant
5	6	6	Grove City	39.8815	-83.093	Pizza Place
5	10	10	Pickerington	39.8963	-82.7732	Pizza Place

4. Results and Discussion

As we can see from the analysis so far, Ohio has sizeable Asian descent immigrant population, and we can see a microcosm of that in the demography of Franklin County. From the census statistics of 2018, for the state of Ohio, we can see the top 5 immigrants – Indian immigrants top the list. In Franklin county we see that Asians make up the third largest grouping. While the distribution within the Asian community is not available for this study, we can safely assume the ratio is similar as at the state level. It is more so prevalent as a significant portion of the Indian population work in the high-tech sector of IT and its associated industry and these industries are primarily located around Columbus. An important factor to keep in mind is the affluence in the cities. Franklin County boasts of 4 of the top 10 Ohio cities in terms of Median Family Income. This and the fact that the cities in the vicinity of Columbus have a significant Indian population which makes it the ideal target for the location of the Indian Restaurant. Coming to the public infrastructure, we saw that the cities are well laid out with freeway connectivity all around

5. Conclusion

Based on the above results and discussion, the recommendation of location of the Indian Restaurant is in any of the three top cities in the Franklin County – New Albany, Upper Arlington, and Dublin.

The stakeholders should also consider other parameters like the real estate cost, density of foot traffic in locating the restaurant.

6. References

1. https://ohio.gov/
2. http://www.co.franklin.oh.us/
3. https://en.wikipedia.org/wiki/List_of_cities_in_Ohio
4. https://en.wikipedia.org/wiki/Ohio
5. https://en.wikipedia.org/wiki/Downtown_Columbus,_Ohio
6. https://www.franklincountyohio.gov/
7. https://en.wikipedia.org/wiki/Franklin_County,_Ohio
8. https://maps2.columbus.gov/arcgis/rest/services/Schemas/PointsOfInterest/MapServer/7/query?where=1%3D1&outFields=*&outSR=4326&f=json
9. https://www.zipdatamaps.com/franklin-oh-county-zipcodes
10. https://www.cleveland.com/datacentral/2017/12/ranking_every_ohio_city_county_2.html
11. https://www.cleveland.com/datacentral/2020/01/every-ohio-city-and-county-ranked-for-median-family-household-income-census-estimates.html
12. https://datausa.io/profile/geo/franklin-county-oh

7. Acknowledgement

1. **Coursera** – For giving me the opportunity to learn the Data Science Discipline
2. **Instructors** – This is an awesome breed – great presentation and keeping me engrossed in the course, providing fantastic hands on labs.
3. **Staff** – for trying to stem the flow of the questions and issues – and trying to resolve each and everyone of them.
4. **IBM** – for providing the top notch cloud environment for me to work on and also providing me the access to Watson Studio and the services in the developer environment – all for free. It is a privilege to work on this environment and learn the latest in Data Science – most of which is the brain child of the great researchers at IBM – Thank You!!

8. Appendix

Further Research to support Conclusion

To check for Radius = 10000 meters - to see the additional Indian Restaurants - But it did not give any additional numbers to influence the conclusion

```
franklin_cities_venues_10000 = getNearbyVenues2(names=franklin_cities['City'],
                                   latitudes=franklin_cities['Latitude'],
                                   longitudes=franklin_cities['Longitude'],
                                         radius=10000
                                  )
```

•••

[97]:
```
print(franklin_cities_venues_10000.shape)
franklin_cities_venues_10000.head()
```

```
(1557, 7)
```

[97]:

	City	Latitude	Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Bexley	39.969238	-82.936864	Franklin Park Conservatory and Botanical Gardens	39.965933	-82.952814	Garden
1	Bexley	39.969238	-82.936864	Franklin Park	39.965362	-82.955221	Park
2	Bexley	39.969238	-82.936864	Johnson's Real Ice Cream	39.957082	-82.925941	Ice Cream Shop
3	Bexley	39.969238	-82.936864	Rubino's Pizza	39.956722	-82.928103	Pizza Place
4	Bexley	39.969238	-82.936864	Jeni's Splendid Ice Creams	39.957373	-82.941965	Ice Cream Shop

[98]:
```
df_restaurants_10000 = franklin_cities_venues_10000[franklin_cities_venues_10000['Venue Category'].str.contains('Indian')]
df_restaurants_10000
```

[98]:

	City	Latitude	Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
291	Dublin	40.099229	-83.114077	Amul India	40.086612	-83.092769	Indian Restaurant
818	Hilliard	40.033814	-83.159611	Amul India	40.086612	-83.092769	Indian Restaurant
1516	Worthington	40.093094	-83.017959	Amul India	40.086612	-83.092769	Indian Restaurant